

Minimum Information Loss Fusion in Distributed Sensor Networks

Daniel Clarke

Centre for Electronic Warfare
Cranfield Defence and Security
The Defence Academy of the United Kingdom
Shrivenham, Oxfordshire, SN6 8LA
Email: d.s.clarke@cranfield.ac.uk

Abstract—A key assumption of distributed data fusion is that individual nodes have no knowledge of the global network topology and use only information which is available locally. This paper considers the weighted exponential product (WEP) rule as a methodology for conservatively fusing estimates with an unknown degree of correlation between them. We provide a preliminary investigation into how the methodology for selecting the mixing parameter can be used to minimize the information loss in the fused covariance as opposed to reducing the Shannon entropy, and hence maximize the information of the fused covariance. Our results suggest that selecting a mixing parameter which minimizes the information loss ensures that information which is exclusive to the estimates from one source is not lost during the fusion process. These results indicate that minimizing the information loss provides a robust technique for selecting the mixing parameter in WEP fusion.

I. INTRODUCTION

In a number of applications ranging from military surveillance through to autonomous robotics, distributed platforms integrated within an ad-hoc network topology are used to sense and estimate the physical state of an unknown number of objects. As such, individual nodes remain naive about global knowledge of the topology or state of the network, having only visibility and access to local information. This provides many benefits, including improved privacy and security of nodes, and a reduced requirement for communications bandwidth. Distributed Data Fusion (DDF) provides a framework for decentralized information sharing and perception within an ad-hoc network topology, where individual nodes utilize only local information [4]. As a result, DDF ensures fast and efficient information sharing between nodes and ensures scalability and flexibility within the network. [14], [3].

Although DDF is theoretically equivalent to a centralized Bayesian fusion, there are a number of practical challenges within its implementation. Estimates obtained from different nodes may have an unknown degree of correlation, and when shared within an ad-hoc network topology, cannot be assumed to be independent. In order to maintain consistent belief, nodes within the network must avoid rumor propagation and subsequent double counting of previously shared information. Integrating redundant information as though it were conditionally independent will result in overconfident estimates of the state uncertainty. While it is possible to track the correlation

or common information between measurements, producing theoretically optimal fused estimates as shown in [9] and [10], this can be prohibitively expensive from both a bandwidth and computational perspective.

Covariance Intersection (CI) [15], [13] provides a sub-optimal algorithm for fusing estimates with an unknown degree of correlation between them, avoiding the independence assumptions required by traditional Bayesian filters. Within CI, a mixing parameter is used to provide a linear combination of the two composite probability density functions (pdfs). This has the effect of inflating the estimates and then applying the Bayes Fusion rule. As a result, the method by which the mixing parameter is selected greatly affects how much information from each composite pdf is used within the fusion process. The original developers selected this parameter by minimizing either the determinant or the trace of the fused covariance, which was shown in [5] to be equivalent to minimizing the Shannon information of the fused density function.

There are many applications which feature distributed networks of sensing nodes observing an unknown number of targets. Within the DDF framework, estimates are shared and fused without knowledge of the common information between them. Furthermore, there are many applications where the sensing nodes have different capabilities, or with different levels of trust (e.g. from false alarm rate). The mixing parameter ω used in covariance intersection offers a framework for weighting the contribution an estimate from any node during the fusion process. While some investigations such as [11] and [12] have focused on methodologies for estimating and exploiting the correlations between estimates stored in different nodes. The primary aim of this paper is to consider the method for optimizing the mixing parameter to ensure the information lost during fusion is minimized.

In this paper, we present preliminary work for understanding how the mixing parameter of CI can be optimally selected. We are specifically interested in the degree of information loss realized by using covariance intersection, and how the mixing parameter can be selected to minimize the information loss from either estimate. This paper is organized as follows: Section II will outline the problem statement, section III will outline the metric for measuring information loss and

section IV will investigate the relation between the CI mixing parameter and information loss.

II. GENERALIZED DESCRIPTION OF THE PROBLEM

In this paper we consider the generalized problem of two sensing nodes observing a single target, fusing shared state estimates using the CI framework. We assume sensor i to be a relatively inaccurate sensor generating many estimates of the target state. We assume sensor j to be a relatively accurate sensor, albeit generating fewer estimates of the target state. These sensors are networked together, where due to the disparate nature of the sensors there is unknown correlation between the measurements.

Traditional centralized fusion frameworks are able to fully recover new information, but are vulnerable to node failures and scale poorly as the number of nodes increases. Furthermore, traditional Bayesian fusion requires that the correlation between estimates is fully known. Distributed Data Fusion (DDF), provides a decentralized framework for information sharing and perception using interconnected sensor systems. Distributed Data Fusion offers a degree of scalability and robustness which is not realized within centralized architectures, overcoming many of the bottlenecks of centralized systems. Furthermore, DDF provides a powerful framework under which estimates can be fused without prior knowledge of the correlation between them [3]. As such, DDF provides an excellent framework with which to approach this problem.

A. Distributed Data Fusion

Consider a distributed system of N independent and heterogeneous sensing platforms observing a target with state $x^k \in \mathbb{R}^{n_x}$, connected within a distributed data fusion architecture. Each sensing platform $i \in 1, \dots, N$ obtains n_y measurements of the target state x_k , in vector y^k at time step k . This produces the likelihood function $p_i(y_i^k | x^k)$. The set of all information is given as $Z_i^k = \{Z_i^{k-1}, y_i^{k-1}\}$, giving the posterior probability density function of $p_i(x^k | Z_i^k)$ for platform i , prior to new information Z_j^k arriving from platform j . Thus, the local fusion of y_i^k is given by the posterior pdf of Bayes rule:

$$p_i(x^k | Z_i^k) \propto p_i(x^{k-1} | Z_i^{k-1}) p_i(y_i^k | x^k) \quad (1)$$

Within a DDF framework, the centralized Bayesian fusion for a pair of platforms (i and j) is given as [6]:

$$p(x^k | Z_i^k \cup Z_j^k) \propto \frac{p_i(x^k | Z_i^k) p_j(x^k | Z_j^k)}{p(x^k | Z_i^k \cap Z_j^k)} \quad (2)$$

The denominator in equation 2 is the common information pdf for the local platform. This term can be calculated by maintaining the probability density between the states of all nodes within the network. However, this is a burden from both a bandwidth and computational perspective. If this term is ignored, it would be double counted, a process often known as rumor propagation and cause the system to become over-certain [6].

Covariance intersection [7] was introduced as a technique which made no assumptions about the independence of the estimates to be combined. It was shown in [8] and [5] that if the estimates are assumed to be Gaussian distributed, the exponential mixture of the two densities is equivalent to the covariance intersection. Therefore, the weighted exponential product (WEP) rule is given as:

$$p_w(x^k | Z_i^k \cup Z_j^k) = \frac{1}{\eta_c} p_i^\omega(x^k | Z_i^k) p_j^{1-\omega}(x^k | Z_j^k) \quad (3)$$

where parameter η_c is a normalizing constant. The mixing parameter $\omega \in [0, 1]$ has the effect of first inflating the component covariances and then applying the standard Bayes fusion rule. The WEP rule is well suited to DDF problems as it is applicable to arbitrary, non-exponential family pdfs and it is guaranteed not to double count the common information ($p(x^k | Z_i^k \cap Z_j^k)$). Unfortunately, the conservative nature of the WEP also means that information which is exclusive to the local platform may be discarded. Therefore, the WEP fusion weight ω must be chosen to ensure that the correct amount of information from the composite pdfs is exchanged.

B. WEP Fusion Metrics

Within the context of equation 3, the optimization of the mixing parameter ω ensures consistency, but has the effect of inflating the component pdfs. It was shown in [5] that minimizing the determinant of the fused covariance is equivalent to minimizing the Shannon entropy of the fused covariance. Effectively, this suggests that minimizing the Shannon entropy effectively finds the fused pdf that contains the most information. However, it was demonstrated in [2] that for a generalized case of the Covariance Intersection fusion rule, Shannon entropy is not a sufficient condition for conservativeness. If we consider that Bayesian data fusion is a methodology for combining information, minimizing the Shannon entropy of the fused pdf will maximize the total information, but may exclude exclusive information contained within one of the other estimates.

Within the scope of this paper, we are specifically interested in minimizing the information loss when state estimates from sources with disparate capabilities are fused. Therefore, we are specifically interested in the derivation of metrics which will minimize the information loss between two estimates, as opposed to maximizing the information.

III. MINIMUM INFORMATION LOSS WEIGHT FUSION

Within CI, minimizing the determinant of the fused covariance maximizes has the effect of summarizing the whole of the relevant information provided by the two samples, essentially maximizing the amount of information. From a similar perspective, we can define the information loss to be a measure of the statistical divergence between the two samples. Thus, within the scope of this paper which seeks to integrate sensors with disparate capabilities where an individual node may not trust the information provided by another node, minimizing the

information loss provides an excellent framework for selecting the mixing parameter ω .

The minimum information loss cost function was introduced as a means for optimizing the CI mixing parameter ω , was introduced in [1]. The information loss was defined as the Kullback-Leibler divergence between the theoretically optimal pdf and the one provided by WEP fusion. Essentially, this selects the value of ω which minimizes how the optimal and WEP fusion can be discriminated from the ideal fusion result.

Within a distributed framework it is not practicable to compute the optimal fusion as this would require maintaining the pdf between the estimates of all nodes within the system. However, within the WEP framework, it is possible to state that if there is no common information, then the optimal fusion pdf simplifies to Naive Bayes (NB) fusion. Therefore, the information of the fusion between estimates of nodes i and j can be defined as the Kullback-Leibler divergence between the naive Bayes fusion p_{NB} and the WEP fusion p_{wep} :

$$L(\omega) = D_{KL}[p_{NB}(x|Z_i \cup Z_j)||p_{wep}(x|Z_i \cup Z_j)] \quad (4)$$

where the parameter ω can be computed by minimizing the function $L(\omega)$ for $\omega \in [0, 1]$ such that:

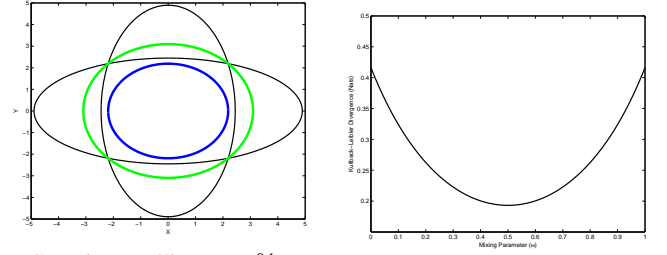
$$\omega^* = \underset{\omega}{\operatorname{argmin}} L(\omega) \quad (5)$$

The principal advantage of this methodology is that the resultant parameter ω^* will be selected to minimize the information loss between the optimal and WEP fusion pdfs. As the Kullback-Leibler divergence measures the statistical difference between two functions, $L(\omega)$ should be a smooth function without any local minima, therefore any numerical optimization technique should be suitable. It should be noted that in the case where the two distributions do not share a significant amount of common information, the solution will tend towards an inflated version of the naive Bayes solution.

IV. SIMULATION RESULTS

The simulated results in this section were implemented using a scenario consisting of two sensors (i and j) observing the location of a single target. We investigate various scenarios which test the information loss rule. In each scenario, we consider how the information loss as defined in equation 4 varies as a function of the mixing parameter ω , the difference between the value of ω selected by minimizing the determinant of the fused covariance; and by minimizing the information loss.

The scenario is set up using two range-bearing sensors observing the target orthogonally. The noise is assumed to be Gaussian such that a sensor measurement for sensor i is given by $y_i = \bar{y}_i + \tilde{y}_i$. For simplicity when we present values in this section, we do so in the x - y space, with the target centered at the origin.



(a) Covariance Ellipses (95% confidence) (b) Kullback-Leibler Divergence

Fig. 1: The fusion of two sensors with similar capabilities. Shown are the input covariances (black), the covariance from naive Bayes Fusion (blue) and covariance from MIL fusion (green). On the right hand side is Information Loss as a function of ω .

A. Ideal Case

In the first case, we show that the minimum information loss fusion will select the mixing parameter ω which is the same as that of the minimizing the determinant of the fused covariance. Each sensor makes a single estimate of the target location which is equal to the target truth location (i.e. no offset). The covariance for sensor i is given as $\mathbf{diag}[1 \ 4]$ and $\mathbf{diag}[4 \ 1]$ for sensor j .

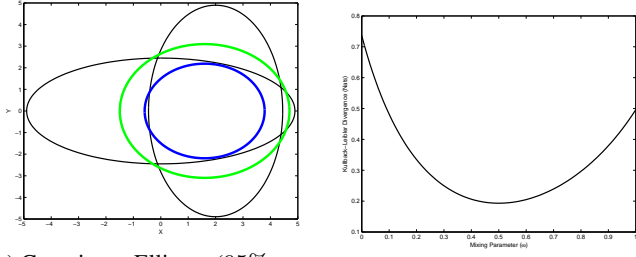
In the ideal example shown in figure 1. Both the CI and minimum information loss fusion rules provide a value of $\omega = 0.5$ for the mixing parameter. Furthermore, there is no difference in information loss between the CI and MIL fusion. This result is fully to be expected; as each covariance is exactly the same (albeit rotated), each will add the same amount of information. Thus, a value of $\omega = 0.5$ will both maximize the amount of information in the fused covariance and similarly minimize the information lost in comparison to a naive Bayes fusion.

B. Offset Covariances

We now consider the same test, but with one of the sensor measurements offset. The offset value is small enough such that it could be considered noise on the sensor measurement. The results, including the ellipses are presented in figure 2. Similar to the previous scenario, both the CI and MIL fusion rules provide a value of $\omega = 0.5$ for the mixing parameter. Furthermore, there is no difference in the information loss between either rule. Again, these results are as expected as the fused pdf is a composite of two similar covariances, with the results indicating that $\omega = 0.5$ both maximizes total information and minimizes the information lost during the WEP fusion.

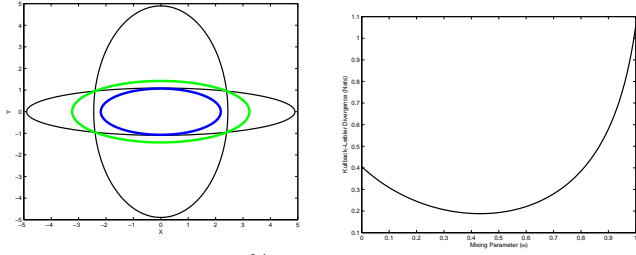
C. Sensors with dissimilar capabilities

For this example we attempt to show the utility of this fusion rule for fusing sensors with disparate capabilities. We use the same scenario as before, however one sensor i uses a covariance of $\mathbf{diag}[1 \ 4]$, while sensor j uses a covariance of $\mathbf{diag}[4 \ 0.2]$. The results shown in figure 3 show the



(a) Covariance Ellipses (95% confidence) (b) Kullback-Leibler Divergence

Fig. 2: The fusion of two sensors with similar capabilities, but with one sensor measurement offset. Shown are the input covariances (black), the covariance from naive Bayes Fusion (blue) and covariance from MIL fusion (green). On the right hand side is Information Loss as a function of ω .



(a) Covariance Ellipses (95% confidence) (b) Kullback-Leibler Divergence

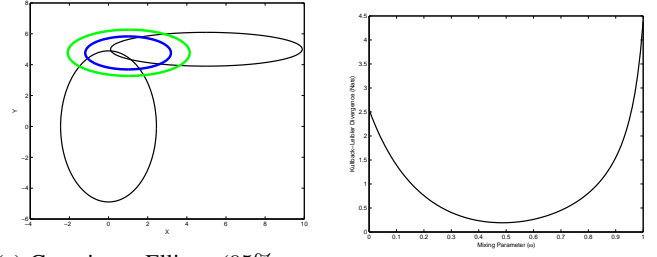
Fig. 3: The fusion of two sensors with dissimilar capabilities. Shown are the input covariances (black), the covariance from naive Bayes Fusion (blue) and covariance from MIL fusion (green). On the right hand side is Information Loss as a function of ω .

covariance ellipses and the relation between information loss and the mixing parameter ω .

Within this example, the two mixing parameters are given as $\omega_{CI} = 0.36$ and $\omega_{MIL} = 0.43$ (i.e. both weighted more towards sensor j). This provides interesting insight into the use of the minimum information loss rule. We observe that in the case of estimates with dissimilar covariances, minimizing the determinant (and thus maximizing the information) of the fused covariance has the effect of discarding some of the information from the estimate with the larger covariance. That is, selecting mixing parameter ω by minimizing the information loss, incorporates more information from both sensors as opposed to favoring one single sensor. While providing a less optimal result, it does highlight the robustness of this technique.

D. Disparate and Offset Measurements

In this next example, we consider a scenario similar to the one presented above. Sensor i uses a covariance of $\text{diag}[1 \ 4]$, while sensor j uses a covariance of $\text{diag}[4 \ 0.2]$. However, in this case, we bias the measurement from sensor j from the center of the measurement, as shown in figure 4a. The method of selecting ω_{MIL} by minimizing the determinant of



(a) Covariance Ellipses (95% confidence) (b) Kullback-Leibler Divergence

Fig. 4: The fusion of two sensors with dissimilar capabilities and with one measurement biased. Shown are the input covariances (black), the covariance from naive Bayes Fusion (blue) and covariance from MIL fusion (green). On the right hand side is Information Loss as a function of ω .

the fused covariance, uses only the composite covariances, giving a value equal to the previous example ($\omega_{CI} = 0.36$). However, the method which minimizes the information loss generates a result which weights the two estimates almost equally ($\omega = 0.49$). This highlights that the fused produced using the minimum information loss rule tends towards an inflated version of the covariance produced by naive Bayes fusion.

E. Representative Scenarios

The results presented above provide an insight into how the minimum information loss fusion rule provides a robust method for selecting the mixing parameter ω . In this subsection, we evaluate this method further within a more operationally relevant scenario. The scenario consists of two sensors: One considered to be a stand-in sensor with relatively low capability continually observing the target; the second sensor is considered a stand-off sensor of high capability providing single measurements against the target. This scenario can be realized in a number of real world applications ranging from car-to-car or car-to-X applications, through to military surveillance.

From a technical perspective, the simulation is set up as follows. Sensor i will measure the location of a target n times, fusing each new result with its prior estimate using naive Bayes fusion. Each measurement from sensor i will be varied around the truth value using a normal random number generator, using the measurement covariance of $\text{diag}[4 \ 4]$. After n measurements, sensor j will observe the target once using a covariance of $\text{diag}[1 \ 0.2]$, sharing this measurement with sensor i for local fusion. For each test of n measurements, we perform a monte-carlo analysis using 1000 iterations.

Figure 5 shows the information loss curves as a function of the mixing parameter ω , for different numbers of sensor measurements (n). Figure 6 shows the values of mixing parameter ω derived using by minimizing the information loss, and minimizing the determinant of the covariance matrix.

These results suggest that by choosing ω to maximize the total information in the fused covariance discards information

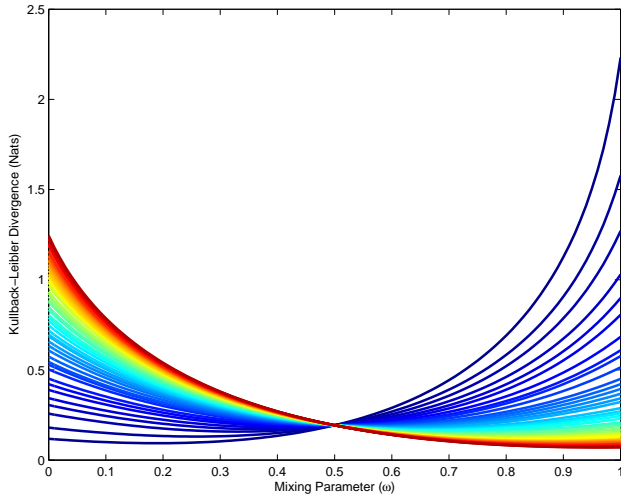


Fig. 5: The information loss for fusing the n measurements low-quality measurements from sensor i , with a single high quality measurement from sensor j as a function of mixing parameter ω . Each information loss curve represents a different value of n , with red starting at $n = 2$ and finishing with Blue at $n = 50$

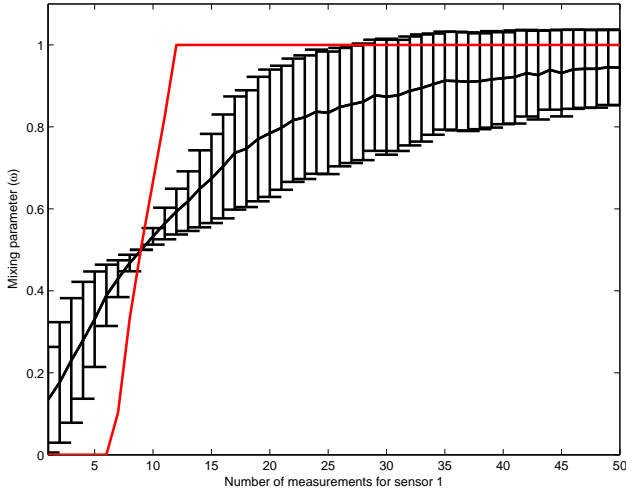


Fig. 6: The values of mixing parameter ω derived by minimizing the determinant of the fused covariance (red), and minimum information loss (black) techniques. The error bars for the minimum information loss values show the variance over the monte-carlo iterations.

from the estimates produced by one sensor. In the case where sensor j has the smaller covariance, its estimate dominates the fused result, whereas in the case where the locally fused estimates of i have a smaller covariance, its estimates dominate the fusion. It is expected that the exact values of n which form this relation will be dependent upon the numerical values given for the covariances of each sensor.

Selecting ω by minimizing the information loss helps to ensure information which is exclusive to any specific node

will remain present in the fusion process. This is an important realization as it shows that even when n becomes very large and sensor i becomes very certain about the target location given its own sensors; the measurement contributed by sensor j will still add information to the fused result.

V. CONCLUSIONS

In this paper we have presented preliminary work to develop an algorithm for fusing sensor measurements from sources with disparate characteristics using the distributed data fusion framework. Specifically, we have investigated a method for selecting the mixing parameter within the weighted exponential product fusion rule which minimizes the information loss between sensor measurements. In cases where the sensor measurement covariances are broadly similar, the value of ω which maximizes the information of the fused covariance, and minimizes the information lost in comparison to naive Bayes is approximately the same. However, our results highlight that for scenarios where sensors have disparate capabilities this rule provides a robust method for selecting the mixing parameter. Furthermore, in scenarios where a single low capability sensor recording many measurements is fused with a single highly capable sensor recording fewer measurements, our results indicate that the minimum information loss fusion rule ensures that information which is exclusive to a particular estimate is guaranteed to be counted.

There are a number of areas which remain unclear with the use of this technique and further research is required. While it is known that the robustness of the minimum information loss rule comes at an expense in terms of optimality, further investigation is required to understand extent and effect distributed data fusion topologies of more than two nodes. Secondly, though it is not investigated here, the minimum information loss rule may provide an interesting framework for fusing sensor measurements with different 'trust' characteristics (i.e. one sensor is known to have a higher false alarm rate). This may provide an interesting baseline from which to consider scenario derived metrics for minimizing the information loss. Finally, experiments should be conducted using physical hardware and multiple ($N > 2$) nodes; thus far the technique has only been tested using statistically well behaved simulated data between two nodes.

ACKNOWLEDGMENT

The Author would like to thank Nisar Ahmed for the initial discussions which stimulated this work.

REFERENCES

- [1] NR Ahmed, JR Schoenberg, and ME Campbell. Fast Weighted Exponential Product Rules for Robust General Multi-Robot Data Fusion. *Robotics: Science and ...*, 2012.
- [2] T Bailey, S Julier, and G Agamennoni. On conservative fusion of information with unknown non-gaussian dependence. *Information Fusion (FUSION ...)*, 2012.
- [3] H Durrant-Whyte, M Stevens, and E Nettleton. Data fusion in decentralised sensing networks. ... *Conference on Information Fusion*, 2001.
- [4] S. Grime and H.F. Durrant-Whyte. Data fusion in decentralized sensor networks. *Control Engineering Practice*, 2(5):849–863, 1994.

- [5] MB Hurley. An information theoretic justification for covariance intersection and its generalization. *Information Fusion, 2002. Proceedings of the Fifth . . .*, 2002.
- [6] SJ Julier, T Bailey, and JK Uhlmann. Using exponential mixture models for suboptimal distributed data fusion. *IEEE Nonlinear Statistical Signal . . .*, 2006.
- [7] SJ Julier and JK Uhlmann. General Decentralized Data Fusion With Covariance Intersection (CI), oct 2001.
- [8] Ronald P. S. Mahler. <title>Optimal/robust distributed data fusion: a unified approach</title>. In Ivan Kadar, editor, *AeroSense 2000*, pages 128–138. International Society for Optics and Photonics, aug 2000.
- [9] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In *Proceedings of the IEEE Conference on Decision and Control*, pages 5492–5498, 2007.
- [10] Lee-Ling Ong, Tim Bailey, Hugh Durrant-Whyte, and Ben Upcroft. Decentralised Particle Filtering for Multiple Target Tracking in Wireless Sensor Networks. In *11th International Conference on Information Fusion*, pages 1–8, 2008.
- [11] S. Reece and S. Roberts. Robust, low-bandwidth, multi-vehicle mapping. In *2005 7th International Conference on Information Fusion*, volume 2, page 8 pp. IEEE, 2005.
- [12] S. J. Julier. Information Fusion, 2009. FUSION '09. 12th International Conference on, 2009.
- [13] Jeffrey K. Uhlmann Simon J. Julier. A Non-divergent Estimation Algorithm in the Presence of Unknown Correlations.
- [14] A.W. Stroupe, M.C. Martin, and T. Balch. Distributed sensor fusion for object position estimation by multi-robot systems. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 2, pages 1092–1098. IEEE, 2001.
- [15] Jeffrey K Uhlmann. General data fusion for estimates with unknown cross covariances. In *Proceedings of SPIE*, volume 2755, page 536, 1996.